Remarks

In the final Office Action dated December 24, 2002, the Examiner rejected claims 1-18 under 35 U.S.C. § 112, first and second paragraphs, and under 35 U.S.C. § 103(a) as being unpatentable over either of Trulson, et al. U.S. Patent No. 5,578,832 (hereinafter Trulson) or Brown, et al. U.S. Patent No. 5,807,522 (hereinafter Brown) in view of Ginestet U.S. Patent No. 6,225,636 (hereninafter Ginestet).

Reconsideration and reexamination of the application is respectively requested in view of the following.

Rejection of Claims 1-18 Under 35 U.S.C. § 112, First Paragraph

Applicants respectfully traverse the rejection of claims 1-18 under 35 U.S.C. § 112, first paragraph. Per the Examiner's request to point to page and line number support for the amendment, the Examiner is respectfully directed to the presently pending application, p. 13, ll. 10-13, which states, "In other words, the correction factors are applied to quantitation or measurement data obtained from microarray images which contain spots having dyes of known or unknown excitation or emission spectra to obtain crosstalk-corrected data." The Examiner is further directed to the presently pending application, p. 6, ll. 17-18, wherein the computer 52 is clearly identified as a quantitation computer (i.e., a computer to perform quantitation in accordance with the present invention).

As such, the term quantitation is fully supported in the present application and the rejection should be withdrawn.

Rejection of Claims 1-18 Under 35 U.S.C. § 112, Second Paragraph

Applicants respectfully traverse the rejection of claims 1-18 under 35 U.S.C. § 112, second paragraph. The Examiner contends that there does not appear to be a discussion or definition of quantitation in the specification. However, the meaning of quantitation would

have been understood by one of ordinary skill in the art at the time of the invention. (A claim term means "what one of ordinary skill in the art at the time of the invention would have understood the term to mean". *Markman v. Westview Instruments, Inc.*, 52 F. 3d 967, 986, 34 USPQ2d 1321, 1335 (Fed. Cir. 1995) (en banc), *aff'd* 517 U.S. 370 (1996).).

In particular, Schena, M., DNA Microarrays: A Practical Approach, April, 1999, page 12 (a copy of which is enclosed for the Examiner's convenience) states, "Quantitation is usually accomplished by superimposing a grid over the microarray image and computing an average intensity value for each microarray element with automated software". Furthermore, Schena, M., Microarray Biochip Technology, January, 2000, page 13 (a copy of which is enclosed for the Examiner's convenience) states, "Data quantitation packages...and other commercial tools perform this function well. Typically, a user-defined gridding pattern is overlaid on the image, and the areas defined by the regular pattern of circles or squares are subjected to data extraction"., and page 179 (a copy of which is enclosed for the Examiner's convenience) states, "After the spot location is determined in the image, a small patch around that location (target region) can be used to quantitate the spot intensity level".

As such, the term quantitation would have been understood by one of ordinary skill in the art at the time of the invention and the rejection should be withdrawn.

Rejection of Claims 1-18 Under 35 U.S.C. § 103

Applicants respectfully traverse the rejection of claims 1-18 under 35 U.S.C. § 103. In particular, independent claims 1 and 10 make it clear that the set of correction factors are applied to quantitation data obtained from generated microarray images containing spots having three or more dyes with excitation or emission spectra to obtain cross-talk corrected data. The Examiner concedes that this feature is neither taught, disclosed or discussed by any of the prior art references of record taken either alone or in combination with one another. Therefore, the Examiner has failed to make a *prima facie* case of obviousness as is required under 35 U.S.C. § 103.

Furthermore, even if the cited references, alone or in combination, resulted in the presently pending invention, the Examiner has failed to provide the motivation to combine the teaching of Trulson or Brown and Ginestet as is required for a *prima facie* case of obviousness under 35 U.S.C. § 103(a). The mere fact that references can be combined or modified, which Applicants do not agree is the case with respect to the cited references, does not render the resultant combination obvious unless the prior art also suggests the desirability of the combination or modification. *See*, M.P.E.P. § 2143.01.

The Examiner has also used impermissible hindsight to combine the teachings of Trulson or Brown and Ginestet to attempt to piece together the Applicants' invention. The teaching or suggestion to make the claimed combination must be found in the prior art, not in the applicant's disclosure. *See*, M.P.E.P. § 2143; *see also*, *In re Dembiczak*, 175 F.3d at 999 ("Combining prior art references without evidence of ... suggestion, teaching, or motivation simply takes the inventor's disclosure as a blueprint for piecing together the prior art to defeat patentability–the essence of hindsight.").

Moreover, "[d]efining the problem in terms of its solution reveals improper hindsight in the selection of the prior art relevant to obviousness." *Ecolochem, Inc. v. Southern Edison Co.*, 227 F.3d 1361, 1372 (Fed. Cir. 2000) (citing *Monarch Knitting Mach. Corp. v. Sulzer Morat Gmbh*, 139 F.3d 877, 880 (Fed. Cir. 1998)). Here, the Ginestet reference cited by the Examiner fails to recognize the problem addressed by the Applicants' claimed invention (i.e., as stated in the presently pending application on page 2, ll. 13-24, "To be accurate, the reader must be able to quantitate the brightness of each microarray spot for each labeled DNA sample used in the experiment".). Ginestet fails to teach applying a set of correction factors to quantitation data as presently claimed. As such, the Examiner has failed to make a *prima facie* case of obviousness as is required under 35 U.S.C. § 103.

Regarding claims 2-9 and 11-18, which depend from claims 1 and 10, Applicants contend that these claims are patentable for at least the same reasons that claims 1 and 10 are patentable. Moreover, Applicants contend these claims recite further limitations,

in addition to the limitations of claims 1 and 10, which render these claims additionally patentable over Trulson or Brown and Ginestet, alone or in combination.

Consequently, in view of the above and in the absence of better art, Applicants' attorney respectfully submits that the application is in condition for allowance which allowance is respectfully submitted.

No fee is believed to be due for the submission of this paper. Please charge any additional fees or credit any overpayments as a result of the filing of this paper to our Deposit Account No. 02-3978 -- a duplicate of the first page of this paper is enclosed for that purpose.

The Examiner is respectfully requested to telephone the undersigned to discuss prompt resolution of any remaining issues necessary to place this case in condition for allowance.

Respectfully submitted,

Mack J. Schermer, et al.

By _____

Thomas W. Saur
Reg. No. 45,075
Attorney/Agent for Applicant

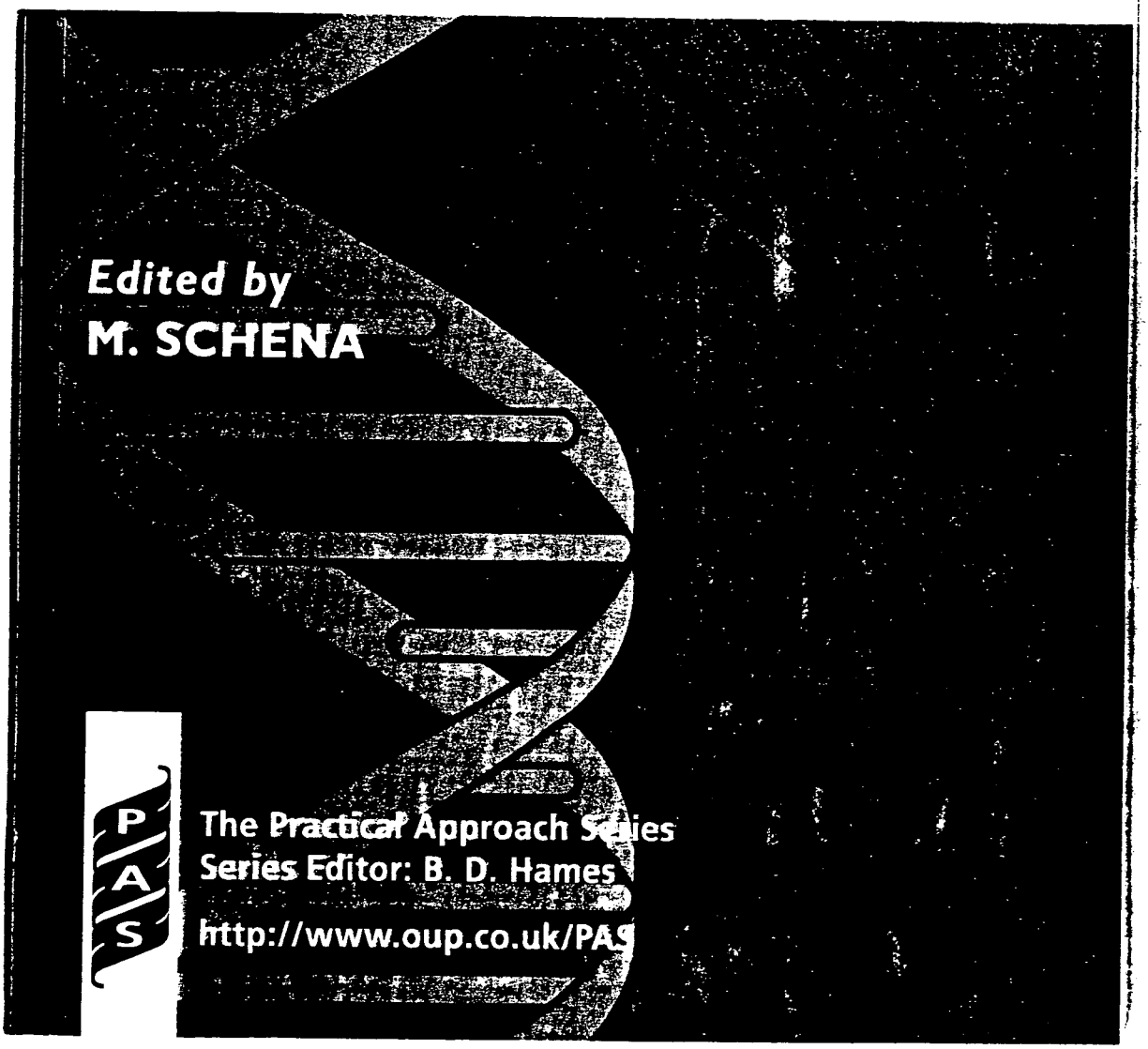
Date: March 13, 2003

BROOKS & KUSHMAN P.C.
1000 Town Center, 22nd Floor
Southfield, MI 48075
Phone: 248-358-4400
Fax: 248-358-3351

Attachments

DNA Microarrays

A PRACTICAL APPROACH



Edited by
M. SCHENA

The Practical Approach Series
Series Editor: B. D. Hames
http://www.oup.co.uk/PAS

OXFORD
UNIVERSITY PRESS
Great Clarendon Street, Oxford OX2 6DP
Oxford University Press is a department of the University of Oxford and furthers the University's aim of excellence in research, scholarship, and education by publishing worldwide in
Oxford New York
Athens Auckland Bangkok Bogotá Buenos Aires Calcutta Cape Town Chennai Dar es Salaam Delhi Florence Hong Kong Istanbul Karachi Kuala Lumpur Madrid Melbourne Mexico City Mumbai Nairobi Paris São Paulo Singapore Taipei Tokyo Toronto Warsaw
and associated companies in Berlin Ibadan
Oxford is a registered trade mark of Oxford University Press
Published in the United States
by Oxford University Press Inc., New York

Reprinted 2000 (three times)

A catalogue record for this book is available from the British Library

Library of Congress Cataloging in Publication Data
DNA microarrays : a practical approach / edited by Mark Schena.
(Practical approach series ; 205)
Includes bibliographical references.
1. DNA microarrays. I. Schena, Mark. II. Series.
QP624.5.D726D63 572.8'65—dc21 99–13842

ISBN 0-19-963777-6 (Hbk)
0-19-963776-8 (Pbk)

Printed in Great Britain by Information Press, Ltd,
Eynsham, Oxon.

Preface

I was 14 years old in the summer of 1977, and it was then that a neighbour first told me about the Stanford Biochemistry Department. Stanford chemistry majors were required to take a class outside their field of immediate study and he opted for a course in biochemistry. Because I had already decided to pursue a career in biochemistry—indeed well before I knew what biochemistry was—it seemed reasonable to ask him about the experience. He said flatly that he enjoyed the course...except for all the biochemistry. And so it was. I was determined to seek out biochemistry and to cross paths with the lauded department that housed the likes of Baldwin, Berg, Hogness, Kaiser, Kornberg, and Lehman.

Enter a spring day in Berkeley in 1983 and a nervous Assistant Professor who had recently completed post-doctoral work at Stanford. Without asking, he volunteered that he had worked for Ron Davis and described him as the 'world's smartest biochemist'. I double checked with Koshland and he confirmed Ron's growing reputation as an enigmatic and brilliant investigator. Later that day and still under the legal drinking age, I had a beer (illegally) and pondered how it was that I would eventually manage to work for Davis.

The UCSF shunt proved unexpectedly prosperous, hitting pay dirt with Yamamoto at an early stage of graduate training. It was in San Francisco, under Yamamoto's stern direction, that I learned how to do science. Marinated in a tonic of Alberts, Bishop, Herskowitz, Pruisner, Varmus, and others, it was difficult *not* to be successful at UCSF.

It was late in 1993 that the daunting complexity of plant transcript factor function led us to some serious pondering and triggered the insane idea that gene expression might best be studied with DNA chips. The notion drew audible laughter from a crowd of 500 in Holland during the summer of 1994. Charging ahead with support from Fodor, Shalon, and Brown, we managed to move the project from infancy to publication in October of 1995.

At present, the microarray field is growing rapidly, fueled internationally by contributors from both academia and industry. The stage is set for a fundamental transformation of biology from a gene science into a genome science. Drawing steadily on expertise from chemistry, engineering, and physics, biochip assays are fast becoming sophisticated and affordable. It is a great privilege to edit a timely book on this subject. Oxford University Press and each of the talented contributors have me wondering, as I have so many times in the past, how it is that I always seem to get much more from science than I deserve.

Stanford
April 1999

M. S.

focal scanners and CCD-based detection systems is that the latter often use continuous wavelength light sources such as arc lamps thereby obviating the need for multiple lasers. Fastidious filtering of emission spectra in CCD-based systems minimizes optical cross-talk between different channels. Detailed descriptions of both confocal scanners and CCD imaging systems are provided in Chapter 2.

Once the fluorescent emission from the microarrays is converted into a digital output by the detection system, the data files (usually 16-bit TIFF) are quantitated and interpreted. Quantitation is usually accomplished by superimposing a grid over the microarray image and computing an average intensity value for each microarray element with automated software. Intensity values can then be converted into biologically relevant outputs such as the number of mRNAs per cell, by comparing the experimental and controls elements present in a given microarray. Quantitative gene expression, genotyping, and other outputs are then correlated with the gene sequences represented in the microarray and higher order relationships such as co-regulation and gene regulatory networks can be identified. Quantitation and data mining software are described in detail in subsequent chapters (see Chapters 2–10).

4. Applications of microarrays

In a manner similar to the polymerase chain reaction (PCR), microarrays have a multitude of applications many of which will develop and evolve over time (21, 25, 29–31). Though the first application of biochips was in gene expression monitoring (9–17), the strategy of using an ordered array of biomolecules on a chip to examine a biochemical sample is generally applicable (*Figure 1*). In addition to gene expression analysis, hybridization-based assays have been used for mutation detection (28, 32–36), polymorphism analysis (2), mapping (37), evolutionary studies (40), and other applications (38, 39). Microarray assays could also be used to monitor the binding of proteins to nucleic acids, small molecules, and other proteins, but these applications have yet to be developed.

Hybridization analysis of genomic DNA can identify single base changes, deletions, and insertions in both coding and non-coding DNA. Hybridization analysis of DNA can also be used to measure the amount of a DNA sequence present, which is important in establishing gene copy number and chromosomal ploidy. In principle, *de novo* sequencing on chips is also feasible but has yet to be demonstrated.

Samples for DNA analysis can obtained either from total genomic DNA or from cloned fragments, by enzymatic incorporation of fluorescent nucleotides. Fluorescent DNA samples can also be obtained by PCR amplification with fluorescent primer pairs. RNA copies of DNA can also be used to examine cloned DNA fragments. RNA probes are usually prepared from cloned DNA by the incorporation of fluorescent nucleotides with RNA polymerase.

Hybridization analysis of RNA can provide information about which genes are expressed in a given sample and at what level. In gene expression applications, fluorescent probes are usually prepared from RNA by enzymatic incorporation of fluorescent nucleotides into complementary DNA (cDNA) by the use of reverse transcriptase. RNA probes for expression monitoring can also be made by linear amplification of cloned cDNA with RNA polymerase (see Chapter 7). In experiments with cDNA microarrays where the hybridization temperatures are sufficient to remove secondary structure in the DNA, a mixture of intact single-stranded molecules (300–3000 nt) provide robust hybridization signals. For assays involving microarrays of oligonucleotides where hybridization temperatures are generally lower, intense hybridization typically requires reducing the size of the molecules in the probe mixture into smaller (50–100 nt) fragments. Nucleic acid size reduction can be accomplished by both chemical (11, 16, 28, 40) and enzymatic means (2).

Unlike DNA and RNA analysis, the use of biochips to explore parallel protein function has proven much more difficult to implement. One shortcoming derives from the fact that many protein–protein interactions, unlike hybridization reactions which involve the interactions of linear sequences, occur with polypeptide surfaces that result from folded, three-dimensional amino acid sequences. The requirement for folded proteins in microarray assays is problematic for several reasons. First, microarray fabrication would have to be implemented in such a way as to maintain the delicate, folded properties of the proteins. The use of harsh chemicals, heat, drying, and the like in protein chip fabrication would compromise the quality of the microarray. Secondly, interactions between folded proteins display a much greater sequence-dependency than hybridization reactions. Sequence-dependency of the interactions inherently complicate reaction kinetics and assay quantitation. Thirdly, the preparation of high quality, fluorescent protein samples remains to be achieved. These and other problems have slowed the implementation of protein microarray technology.

5. Chips and pharmacogenomics

One exciting application of microarray technology is in the area of pharmacogenomics, a new field in biomedicine focused at the interface between pharmacology and genomics (*Table 4*). Pharmacology is the branch of science that endeavours to understand the preparation, use, and effects of drugs. Genomics endeavours to provide the complete sequence of genomes, and a complete knowledge of genes and gene functions. The two disciplines are complementary and synergistic in several respects (*Table 4*).

The complete sequence of all of the genes in a genome provides, by definition, a sequence of all of the gene products encoded by the genome. Because most drugs act at the protein level to disrupt or alter protein function, a complete genomic sequence provides all of the potential drug targets for a given

BioTechniques®
BOOKS

Microarray Biochip Technology

Edited by
Mark Schena

Mark Schena
TeleChem/arrayit.com
524 E. Weddell Drive
Suite 3
Sunnyvale, CA 94089-2115, USA

Library of Congress Cataloging-in-Publication Data

Schena, Mark.
Microarray biochip technology / Mark Schena.
p. cm.
Includes bibliographical references and index.
ISBN 1-881299-37-6
1. Biochips. 2. DNA microarrays. I. Title.

R857.B5 S35 2000
660'.65--dc21

99-088465

ISBN 1-881299-37-6

Printed in the United States of America

9 8 7 6 5 4 3

Eaton Publishing
BioTechniques Books Division
154 E. Central Street
Natick, MA 01760
www.BioTechniques.com

Francis W. Eaton: *Publisher and President*
Stephen Weaver: *Director and Editor-in-Chief*
Christine McAndrews: *Managing Editor*
Sandy Lamont: *Production Manager*
Paul Haje: *Cover Designer*

CONTRIBUTORS

Karl Adler
NEN Life Science Products, Inc.
Boston, MA, USA

David Barker
Molecular Dynamics, Inc.
Sunnyvale, CA, USA

Trent Basarsky
Axon Instruments, Inc.
Foster City, CA, USA

Kate Bechtol
Molecular Dynamics, Inc.
Sunnyvale, CA, USA

Jeffrey Broadbent
NEN Life Science Products, Inc.
Boston, MA, USA

Jing-Ying Chen
BioDiscovery, Inc.
Los Angeles, CA, USA

Ronald W. Davis
Department of Biochemistry
Beckman Center
Stanford University
Medical Center
Stanford, CA, USA

Suzanne Dee
Affymetrix, Inc.
Santa Clara, CA, USA

David Englert
Packard Instrument Company
Meriden, CT, USA

Elisabeth Evertsz
Incyte Pharmaceuticals, Inc.
Palo Alto, CA, USA

Russell Garlick
NEN Life Science Products, Inc.
Boston, MA, USA

Robert Gupta
Incyte Pharmaceuticals, Inc.
Palo Alto, CA, USA

Paul Haje
arrayit.com
Sunnyvale, CA, USA

David Hanzel
Molecular Dynamics, Inc.
Sunnyvale, CA, USA

John Hartwell
Gene Logic, Inc.
Gaithersburg, MD, USA

Glenn Hoke
Gene Logic, Inc.
Gaithersburg, MD, USA

Richard Joseph
NEN Life Science Products, Inc.
Boston, MA, USA

Peter Kalocsai
BioDiscovery, Inc.
Los Angeles, CA, USA

Anis Khimani
NEN Life Science Products, Inc.
Boston, MA, USA

Jeffrey Killian
NEN Life Science Products, Inc.
Boston, MA, USA

Ants Kurg
Institute of Molecular and
Cell Biology
University of Tartu
Estonian Biocentre, Tartu, Estonia

Elin Lõhmussaar
Institute of Molecular and
Cell Biology
University of Tartu
Estonian Biocentre, Tartu, Estonia

Pixel Size

The spatial resolution of detection instruments is typically given as the pixel size, the physical "bin" in which a single datum is acquired. A reasonable guideline is that the pixel size should be approximately 1/10 the diameter of each spot or feature on the chip. Microarrays containing 100-μm features require fluorescent detectors with approximately 10-μm pixel size, 50-μm features require approximately 5-μm pixels, and so forth. Most charge-coupled device (CCD)-based detection instruments are equipped with 1-megapixel cameras that contain 1000 × 1000 chips for a pixel size of approximately 20 μm when imaging a 20 × 20-mm area. At the moment, fluorescent scanners are probably best suited for imaging fairly large arrays, and CCD-based systems are most compatible with arrays ≤1.0 cm^2.

Photobleaching

Microarray dyes emit fluorescent photons in the presence of excitation light. If the light intensity is excessive, however, the incoming photons can damage the dyes and reduce the fluorescent readout of the microarray during successive scans. Photobleaching is particularly problematic if users excite a portion of the array and then interrupt the imaging to adjust laser or PMT settings before scanning the remaining array area. Photobleaching can greatly degrade array data when portions of arrays are selectively exposed to excitation light. Most commercial scanners use 5- to 35-mW light sources with scanning speeds of 5 to 25 lines per second. More powerful light sources or longer pixel dwell times can lead to considerable photobleaching. A guideline is that photobleaching should be <1% per imaging cycle.

It should be noted that dyes differ considerably in their photostabilities, with fluorescein isothiocyanate (FITC) being more susceptible to photobleaching than the Cy dyes. Differential photostability can be particularly problematic in experiments involving multiple fluors and ratio measurements in that photobleaching in one channel (but not the other) can lead to errors in quantitation. Photobleaching should be minimized as much as possible.

Cross-Talk

Cross-talk refers to emitted light from one channel being detected in another channel. This optical "contamination" primarily occurs because emission spectra are typically rather broad, with as much as 10% of the emitted signal observable at wavelengths 100 nm longer than the peak emission wavelength. In dual-labeling and detection experiments, fluorescence intensity from the Cy3 channel, for example, can contaminate the Cy5 channel if care is not taken to minimize cross-talk. As the Stokes shift dictates, the emission wavelength is always longer than the absorbance wavelength, and therefore Cy3 cross-talk into the Cy5 channel is a practical problem but not vice versa. There are many ways to minimize cross-talk, although the most common and least expensive involves the use of emission filters that reject light outside the desired wavelengths. Optical cross-talk should be kept to <0.1%/channel, with lesser amounts desired for high-sensitivity applications such as gene expression analysis.

DATA FILES AND MINING

Microarray experiments generate an enormous data stream, placing bioinformatics at the core of the microarray industry. Attention must be paid to data output and modeling. The most common graphical output format from fluorescence imaging instruments is the 16-bit tiff (.tif) file. The 16-bit format allows intensity readings from 0 to 65 536 (2^{16}) to be recorded, thereby permitting a theoretical dynamic range in excess of four orders of magnitude. Instrument makers should probably embrace the 16-bit tiff file format as the format of choice for the present time, with a 16-bit bitmap (.bmp) format as a common alternative.

Output graphical files in 16-bit format are, in essence, two-dimensional number sets with values ranging from 0 to 65 536. The intensity values corresponding to the regular pattern of each microarray element are generally higher than the background fluorescence values corresponding to the area between the spots. Extracting the data is basically a task of summing the pixel values of each spot and dividing by the total quantified area. Tiff files allow a single microarray image to be mined with multiple quantitation packages. A 16-bit output also allows data sets from different instruments or laboratories to be compared with reasonable precision. Output files of 12 and 8 bits reduce the storage space required for the images, but these formats only allow 4096- and 256-fold theoretical dynamic ranges, which is insufficient for gene expression applications.

Data quantitation packages such as QuantArray™ from GSI Lumonics, ImaGene™ from BioDiscovery (Los Angeles, CA, USA), the software on the Axon GenePix™ system, and other commercial tools perform this function well. Typically, a user-defined gridding pattern is overlaid on the image, and the areas defined by the regular pattern of circles or squares are subjected to data extraction. The computational theory underlying data quantitation is presented in detail in Chapter 8.

Once values are quantified from the microarray, the data are commonly displayed in a variety of ways. One common depiction is known as a scatter plot, in which all the data points from a two-color experiment are plotted as a function of ratio and signal intensity, with those ratios greater than 1.0 plotted above the diagonal and those with ratios less than 1.0 plotted below the diagonal. The scatter plot provides a means of viewing very large datasets such as those involving expression profiles of thousands of genes. A mouse click on a particular data point usually provides information about the gene sequence present at that position on the microar-

processing approach. The computer itself does not have any visual capabilities to "see" the spots, rather, it provides tools to allow users to tell the computer the location of each signal spot in the image. Typically, a grid frame is given that the user can place manually on the image and then manipulate to fit the spots in the image. Because the spots in the image may not be spaced evenly, the user may need to adjust the grid lines individually to align with the arrayed spots. The size of each circle may also need manual adjustment to fit the size of a particular spot. To conduct an accurate measurement, the manual spot finding method is prohibitively time-consuming and labor-intensive for microarray images with

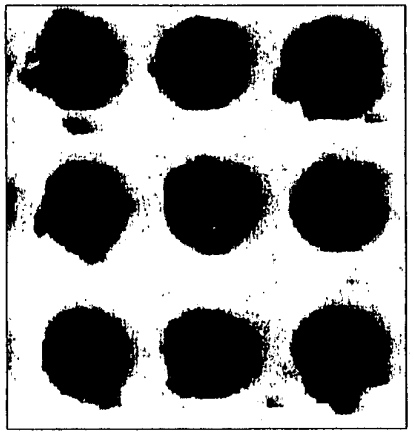


Figure 6. Irregularity of the spot shape. The shape of the spot may deviate from a perfect circle, and the intensity inside the spot may vary considerably.

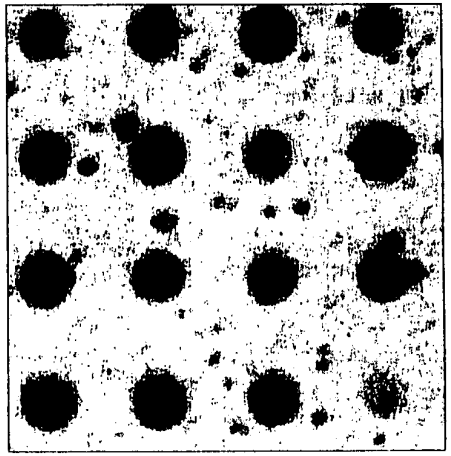


Figure 7. Contamination in microarray images. Punctate contamination is visible both in the background and inside of the spot signal regions.

thousands of spots. Thus, considerable imprecision may be introduced due to human errors, particularly with arrays having irregular spacing between the spots and large variation in spot sizes.

Semiautomatic spot finding. The semiautomatic method requires some level of user interaction. This approach typically uses algorithms for automatically adjusting the location of the grid lines, or individual grid points after the user has specified the approximate location of the grid. What the user needs to do is to tell the program where the outline of the grid is in the image. For example, the user may need to put down a grid and adjust the size of it to fit on the array of the spots, or to tell the program the location of the corners of the subgrids in the images. Then the spot finding algorithm adjusts the location of the grid lines, or grid points, to locate the arrayed spots in the image. User interface tools are usually provided by the software to allow for manual adjustment of the grid points if the automatic spot finding method has not correctly identified each spot.

This approach could potentially offer great time savings over the manual spot finding method since the user needs only to identify a few points in the image and make minor adjustments to a few spot locations if required. The key issue for spot finding algorithms here is to identify the spots correctly even at very low levels of intensity. In addition, the algorithm must deal effectively with two opposing criteria. First, due to variation in spot position, as described earlier, the algorithm must tolerate a certain degree of irregularity in the spot spacing. At the same time, the algorithm must not be "distracted" by contaminants that could be adjacent to a true arrayed spot. Such an algorithm has been developed and implemented in BioDiscovery s ImaGene™ software.

Automatic spot finding. The ultimate goal of array image processing is to build an automatic system that utilizes advanced computer vision algorithms, to find the spots without the need for any human intervention. Such a method would greatly reduce human intervention, minimize the potential for human error, and offer a great deal of consistency in the quality of data. One such system, called AutoGene™, has been developed by BioDiscovery. This tool only requires the user to specify the configuration of the array (i.e., number of rows and columns), and will automatically search the image for the correct grid position. In many tests, the software has been able to locate grids as accurately as those obtained by manual placement of the grid.

Methods of spatial segmentation of signal and background pixels. After the spot location is determined in the image, a small patch around that location (target region) can be used to quantitate the spot intensity level. The next step is to determine which pixels in the target region are signal and which are background. This operation is called signal or image segmentation in computer vision terminology. At this stage, size and shape irregularities of the spots and any contamination problem in the images are the major concerns to the algorithm design. A number of methods have been developed with different levels of sophistication. Their advantages and disadvantages are described below.

ARRAY IMAGE ANALYSIS II: DATA QUANTIFICATION AND QUALITY MEASUREMENT

Methods of Quantitation

On a single microarray chip, the expression levels of many genes are in parallel. Under the proper conditions, the total fluorescence intensity of a spot is proportional to the expression level of a gene. These conditions are as follows:

1. The preparation of the probe cDNA (through reverse transcription of the extracted mRNA) solution is performed such that the probe cDNA concentration in the solution is proportional to the mRNA in the tissue.
2. The hybridization experiment is performed such that the amount of cDNA binding to each spot is proportional to the partial concentration of each cDNA species in the probe solution.
3. The amount of cDNA target deposited at each spot during the chip fabrication is constant and in approximately 10-fold excess relative to the most abundant species in the probe solution.
4. There is no contamination on the spots.
5. The signal pixels are correctly identified by the image processing.

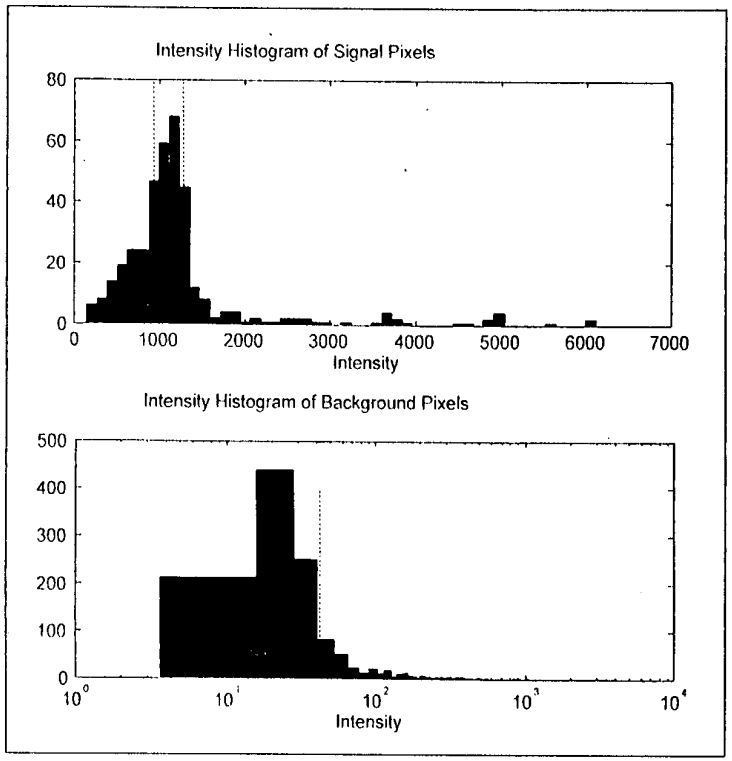


Figure 9. Trimmed measurement method for signal pixel identification. Top, plot of the intensity histogram of the pixels inside the target circle. Thirty percent of the pixels are trimmed off at the lower intensity side (left of the left vertical line). Twenty percent of the pixels are trimmed off the high-intensity side (right of the right vertical line). Bottom, plot of the histogram of the pixels outside the target circle. Thirty percent of pixels are trimmed off at the high-intensity side of the histogram (right of the vertical line). The abscissa is a log scale to show the detail on the lower intensity side. The mean intensities are 1100 for signal and 20 for background. They differ by 32% and -29% from the true value. **(See color plate A15.)**

In the following discussion, we assume that conditions 1 and 2 are satisfied. Whether these two conditions are truly satisfied is determined through the design of the experiments. For the quantitation measurements, the more closely conditions 1 to 5 are followed the better. Often, conditions 3, 4, and 5 are violated to varying degrees. The DNA concentrations in the spotting procedure may vary from time to time and spot to spot. Higher or lower concentrations may result in altered signals. When adjacent spots overlap, the signal intensity corresponding to the contaminated region is not measurable. The image processing may not correctly identify all the signal pixels; thus, the quantification methods should be designed to address these problems. The commonly used methods are total, mean, median, mode, volume, intensity ratio, and the correlation ratio across two channels. The underlying principle for judging which one is the best method is based on how well each of these measurements correlates to the amount of the DNA probe hybridized to each spot location.

Total. The total signal intensity is the sum of the intensity values of all the pixels in the signal region. As has been indicated above, total intensity is sensitive to variations in the amount of DNA deposited on the surface, the existence of contamination, and anomalies in the image processing operation. Because these problems occur frequently, the *total* may not be an accurate measurement.

Mean. The mean signal intensity is the average intensity of the signal pixels. This method has certain advantages over the *total.* Very often the spot size correlates with the samples and pins used in the arraying step. Measuring the mean will reduce the error caused by the variation of the amount of DNA deposited on the spot. With advanced image processing allowing for accurate segmentation of contaminated pixels, the mean is perhaps the best measurement method.

Median. The median of the signal intensity is the intensity value that splits the distribution of the signal pixels such that the number of pixels above the median intensity is the same as the number below the median intensity. The median is a landmark in the intensity distribution profile. The advantage of choosing this landmark as the measurement derives from the resistance of the median value to outliers. As discussed in the last section, contamination and problems in the image processing operation introduce outliers in the sample of signal pixels. The mean measurement is very vulnerable to these outliers. When the distribution